A FAST AND ACCURATE TRACKING ALGORITHM OF LEFT VENTRICLES IN 3D ECHOCARDIOGRAPHY

Lin Yang[§†‡], *Bogdan Georgescu*[§], *Yefeng Zheng*[§], *David J. Foran*[‡], *Dorin Comaniciu*[§]

[†]ECE, Rutgers University; [‡]CINJ, UMDNJ-RWJMS
Piscataway, NJ 08854

[§]Siemens Corporate Research
Princeton, NJ 08540

ABSTRACT

Tracking of left ventricles in 3D echocardiography is a challenging topic because of the poor quality of ultrasound images and the speed consideration. In this paper, a fast and accurate learning based 3D tracking algorithm is presented. A novel one-step forward prediction is proposed to generate the motion prior using motion manifold learning. Collaborative trackers are introduced to achieve both temporal consistence and tracking robustness. The algorithm is completely automatic and computationally efficient. The mean point-to-mesh error of our algorithm is 1.28 mm. It requires less than 1.5 seconds to process a 3D volume (160 × 148 × 208 voxels).

Index Terms— Tracking, Ultrasound, Left Ventricles

1. INTRODUCTION

The 3D echocardiography (ultrasound of the heart) is one of the most widely used diagnostic tools in modern imaging modalities for visualizing cardiac structure and diagnosing cardiovascular disease. There are several advantages of using 3D ultrasound over other imaging modalities, like CT and MRI: 1) Ultrasound is much cheaper than CT and MRI and it is more convenient to use, e.g., hand-carried ultrasound equipment is widely used for routine diagnosis; 2) Ultrasound is noninvasive, which does not produce ionizing radiation or require contrast agents. However, ultrasound imaging normally provides noisy images with poor object boundaries.

Recently, the automatic segmentation and tracking of heart ventricles have received considerable attentions [1, 2, 3, 4]. Among these applications, the tracking of left ventricles (LV) have attracted particular interests, because it provides clinical significance for doctors to detect the coronary artery disease and evaluate acute myocardial infractions. However, tracking LV in 3D echocardiography is still a challenging problem. Widely used 2D tracking algorithms may bring computational problems for a 3D application. The ultrasound image also has relatively low qualities than natural image sequences, which may further bring more frequent tracking failures.

Recently, the idea of utilizing detection for tracking to achieve the robustness in noisy environment is proven to be quite effective. Tracking by detection does not accumulate errors from previous frames and can therefore avoid template drifting. However, it still has two major problems in 3D boundary tracking: 1) The boundary classifiers are sensitive to initial positions [5]. In order to achieve accurate boundary tracking results, good initializations have to be provided. 2) Tracking by detection applies universal description of the objects without considering the temporal relationships. This leads to the temporal inconsistence between adjacent frames.

To address the limitations of the previous work, we propose a new method and make the following contributions:

- A novel one-step forward prediction using motion manifold learning. The learned motion modes provide required good initialization for the boundary classifiers.
- A collaborative 3D template tracker is introduced to erase the temporal inconsistence introduced by detection tracker.
- The algorithm we proposed is fast and accurate. It took less than **1.5 seconds** to process a 3D volume containing 160 × 148 × 208 voxels. The final average *point-to-mesh* error (PTM) we obtained is **1.28 mm**. Considering the resolution of the test dataset, we obtained **subvoxel** tracking accuracy.

Section 2 illustrates the learning procedure. Section 3 describes the tracking algorithm. Section 4 provides the experimental results and section 5 concludes the paper.

2. LEARNING

Because the motion of LV is close to periodic, motion priors play a key role in improving the tracking accuracy. Multiple motion modes are learned using manifold learning and hierarchical K-means. Since the 3D ultrasound has relatively low image quality, learning based 3D active shape model (ASM) [5] is used to achieve the robust 3D boundary tracking. Marginal space learning (MSL) [1] and probability boosting tree (PBT) [6] are applied to train an ED detector to automatically locate the pose of LV in the first frame. Two boundary classifiers are also learned to segment the LV in each frame based on MSL and PBT.

2.1. Learning the Motion Modes on The Manifold

Before motion manifold learning, the first step is the generalized procrustes analysis (GPA). All annotated 3D LV shapes

978-1-4244-2003-2/08/$25.00 ©2008 IEEE

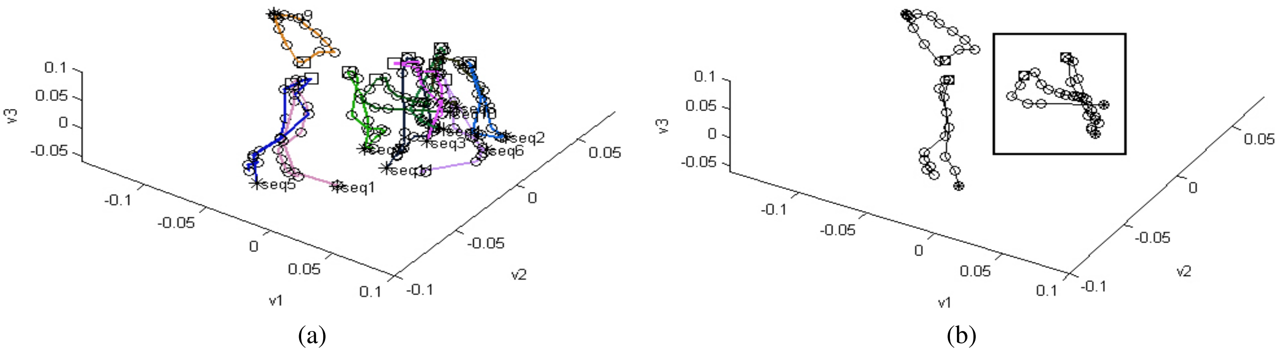

Fig. 1. Manifold embedding of LV motions (a) The 11 LV motion sequences represented with different colors. (b) T clustering results on the embedded low dimensional subspace. The star represents the end diastolic (ED) phase and square denotes the end systolic (ES) phase.

in one training motion sequence are stacked together and temporally resampled to form a motion vector with same dimensionality. The 4D generalized procrustes analysis (GPA) is used to align these motion vectors to remove the translation, rotation and scaling. The shape difference and motion patterns are still preserved. After the 4D GPA, these aligned motion vectors are decomposed into 3D shapes. All the following learning operations are performed on these aligned 3D LV shape vectors.

Given the fact that the actual number of constraints that control the LV motion are less than its original dimensionality, the aligned 3D LV shape vectors are expect to lie on a low dimensional manifold, where geodesic distance has to be used to measure the similarities. Given a set of 3D shape vectors $S = \{\mathbf{s}_1, ..., \mathbf{s}_i, ..., \mathbf{s}_n\}$ where $\mathbf{s}_i \in R^d$, there exists a mapping T which can represent $\mathbf{s}_i$ in the low dimension as

$$\mathbf{s}_i = T(\mathbf{v}_i) + \mathbf{u}_i \quad i = 1, 2, ..., n \tag{1}$$

where $\mathbf{u}_i \in R^d$ is the sampling noise and and $\mathbf{v}_i \in R^{d^{'}}$ denotes the representation of the original i-th shape $\mathbf{s}_i$ in the low-dimensional subspace with dimensionality $d^{'}$.

Unsupervised manifold learning is capable of discovering the nonlinear degrees of freedom that underlie the manifold. We apply ISOMAP [7] to embed the nonlinear manifold into a low dimensional subspace. We first determine the neighbors of each vector $\mathbf{s}_i$ in the original space R^d and connect them to form a weighted graph G. The weights are calculated based on the Euclidean distance between each connected pairs of vectors. We then calculate the shortest distance in the graph G, $d_G(i, j)$, between pairs of vectors $\mathbf{m}_i$ and $\mathbf{m}_j$. The final step is to apply the standard multiple dimensional scaling (MDS) to the matrix of graph distance $M = \{d_G(i, j)\}$. In this way, the ISOMAP applies a linear MDS on the local patch but preserve the geometric distance globally.

Figure 1a shows the 11 LV motion sequences on the embedded low dimensional subspace. It can be observed that the motion of LV roughly form a circle through manifold learning, which proves that the LV motion is pseudo-periodic. Given all the motion cycles shown on the reduced subspace. We applied a hierarchical K-means to learn the motion modes. The clustering results of 11 motion sequences are shown in Figure 1b. The two clustered motion modes (shown in the black rectangle in Figure 1b) represent two complete different motion trajectories which start from similar ED shapes. Each motion mode is a weighted sum of all sequences that are clustered into the same group. The weights are proportional to their Euclidean distance to the cluster center on the reduced subspace. Geodesic distance in the original manifold is modeled by Euclidean distance on the embedded low dimensional subspace.

2.2. Learning The Detector and Boundary Classifiers

Discriminative learning based approaches have proven to be efficient and robust for 2D object detection. In these methods, the object is found by scanning the classifier over an exhaustive range of possible locations, orientations, and scales in an image. However, it is challenging to extend them to 3D problems since the number of hypotheses increase exponentially with respect to the dimensionality of the parameter space. The idea for marginal space learning (MSL) [1] is not to learn a classifier directly in the full similarity transformation space, but incrementally learn classifiers on projected marginal spaces. As the dimensionality increases, the valid (positive) space region becomes more restricted by previous marginal space classifiers. In our case, we split the estimation into three problems: position estimation, position-orientation estimation, and full similarity transformation estimation. MSL can reduce the number of testing hypotheses by several orders of magnitude.

In order to achieve the boundary tracking, Active shape models (ASM) [5] are used in our algorithm. The original ASM does not work in our application due to the complex background and weak edges. Learning based methods can exploit more image evidences to achieve robust boundary classification. We train an ED detector using MSL and two boundary classifiers (one for LV motion close to the ED phase and

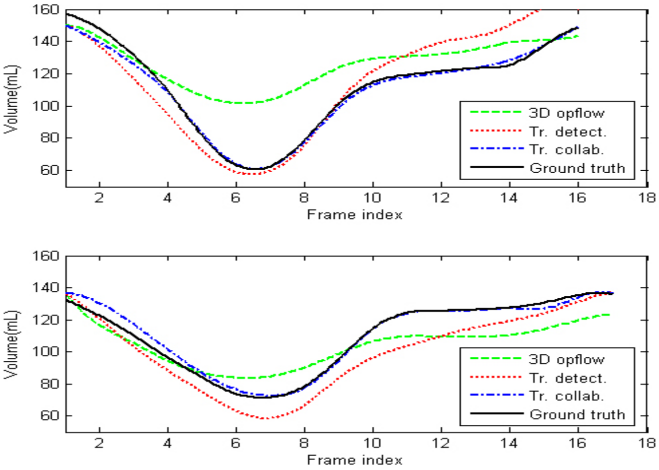

Fig. 2. Two volume-time curves which demonstrate a whole cardiac cycle. The 3D opflow represents the tracking result using 3D optical flow. Tr. detect. represents the tracking by detection and Tr. collab. denotes the results using our algorithm based on collaborative trackers.

the other for the ES) based on probability boosting tree [6] . The ED detector is used to locate the LV and the boundary classifiers are used to segment the 3D LV boundary.

3. TRACKING

Given a testing LV motion sequence, the tracking is initialized from an automatic detection and segmentation of LV in the ED frame using the learned detector and boundary classifiers. At time t, registration based reverse mapping and one-step forward prediction is used to generate the motion prior for $t+1$. Started from the motion prior, two collaborative trackers are used to track the LV in each frame.

3.1. Tracking Initialization

Given the first frame in the LV motion, all positions, orientations and scalings are scanned by trained detector and the first 100 candidates are kept. The final similarity transformation is obtained by simply average the 100 candidates. After the similarity transformation between the mean LV shape and the testing object is found, we put the registered LV mean shape as the initial position for the boundary classifiers. We use marginal space learning (MSL) to detect LV in the first frame. For more details about MSL, we refer readers to [1].

3.2. Registration Based Reverse Mapping and One-Step Forward Prediction

Given the LV shape at time t, in order to obtain the motion prior for $t+1$, we need to map the current LV shape in the real world coordinate system to the leaned multiple motion modes. Thin plate spline (TPS) transformation [8] is applied to perform this mapping. TPS is a nonrigid transformation between two point sets. Affine transformation has proven to be a special case of TPS. Given two 3D point sets, the TPS is estimated by minimizing

$$E_{tps}(T) = \sum_{i} \|w_i - T(v_i)\|^2 + \lambda f(T). \qquad (2)$$

with w_i denote the 3D boundary point on the learned motion modes and v_i denote those on the boundary of LV in the testing motion.

Given the current LV boundary in a testing sequence, the one-step forward prediction is calculated iteratively using the J-th motion mode which minimize the previous t *accumulated* TPS registration errors

$$J = \arg\min_{j} \sum_{i=0}^{t-1} E_{tps}(x_t, m_j), j = 1, 2, ..., N \qquad (3)$$

where x_t is the current LV boundary and m_j represents the corresponding 3D shape of the j-th motion mode. The N is equal to the number of motion modes. Notice that there exists motion mode change during the prediction, where it starts from one motion mode but jumps to another. This corresponds to the LV motion which starts from a similar ED shape with one learned motion mode, but has a motion trajectory close to another. Using the *accumulated* TPS registration error based one-step forward prediction, the algorithm provides accurate motion prior for boundary classifiers.

3.3. Collaborative Trackers

Given the shape prior learned using one-step forward prediction, for each point and its ±12 mm range on the normal directions, the learned boundary classifiers are used to move each point to the optimal position where the estimated boundary probability is maximized. The ED boundary classifier is used when the frame index is close to ED and the ES boundary classifier is used when it is close to ES.

In order to compensate the drawbacks of detection tracker we mentioned in the introduction, the 3D template tracker is also applied. Given $x_t = (x, y, z)^T$ to be the pixel coordinates of a boundary point, we can construct a template around the neighborhood of x_t, $T(x_t)$ ($13 \times 13 \times 13$ cube in our case). Let $G(x_t, \lambda)$ denotes the allowed transformation of the template $T(x_t)$, the goal is to search best transformation parameters which minimize the error between $T(x_t)$ and $G(x_t, \lambda)$.

$$\lambda = \arg\min_{\lambda} \sum_{x_t \in T} [G(x_t, \lambda) - T(x_t)]^2 . \qquad (4)$$

Although the template matching algorithm is not robust and only works under the assumption of small inter-frame motions, it respects temporal consistence. In each frame we update the template using the previous collaborative tracking result, which fuse both the detection tracking and template tracking. Because the global motion prior is enforced, this updating scheme can help template tracker to recover from the tracking failures.

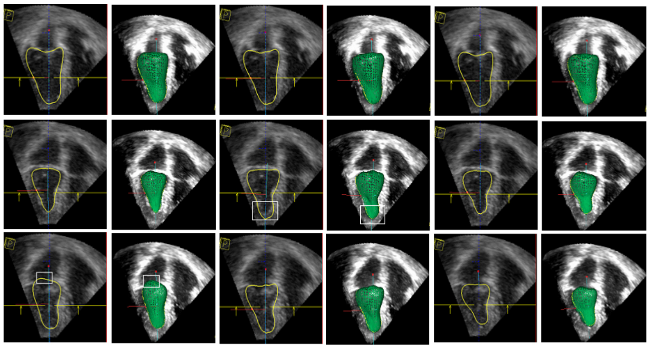

Fig. 3. A comparative tracking results of a testing LV motion sequence with 12 frames. The first two columns are the tracking by detection, the 3rd and 4th columns are the 3D optical flow and the last two columns are the results of our proposed algorithm. The rows correspond to frame index 1, 6 and 8.

The data fusion of two tracking results is obtained by defining prior distribution of detection tracker and template tracker. The detection tracker is assigned more weights around the ED and ES phases while the template tracker is weighed more between the ED and ES phases based on the knowledge of experts.

4. EXPERIMENTAL RESULTS

We collect 67 annotated 3D ultrasound LV motion sequences. Each 4D $(x, y, z + t)$ motion sequence contains 11-25 3D frames. In total we have 1143 3D ultrasound volumetric data. Our dataset is much larger than many reports listed in the literature, e.g. 29 cases with 482 3D frames in [3], 21 cases with about 400 3D frames in [9] and 22 cases with 328 3D frames in [10]. The imaging protocols are heterogeneous with different capture ranges and resolutions. The dimensionality of 27 sequences is $160 \times 144 \times 208$ and the other 40 sequences is $160 \times 144 \times 128$. The x, y and z resolution ranges are $[1.24\ 1.42]$, $[1.34\ 1.42]$ and $[0.85\ 0.90]$ mm. In our experiments, we randomly select 36 sequences for training and the rest is used for testing.

The accuracy is measured by the *point-to-mesh* (PTM) error, e_{ptm}. All 3D points on each frame of the testing sequence are projected onto the corresponding annotated boundary. The projection distance is recorded as e_{ptm}. For a perfect tracking, the e_{ptm} should be equal to zero for each 3D frame. The final mean e_{ptm} we obtained is 1.28 ± 1.11 mm with 80% of the errors below 1.47 mm. Considering the range of resolution in the testset, we actually obtained subvoxel tracking accuracy.

The volume-time curve of LV is an important diagnosis term to evaluate the health condition of the heart. In Figure 2 we show two volume-time curves of the ground-truth annotations, the tracking results using our algorithm and two comparative tracking methods. It is obvious that our algorithm provides the most accurate volume-time functions.

In Figure 3, tracking by detection produces leakage errors in the mitral valve region (white rectangles in columns 1 and 2). The 3D optical flow algorithm fail to produce enough shrinkage in the apex of the heart (white rectangles in columns 3 and 4). Using our proposed algorithm (columns 5 and 6), none of the errors are observed.

One of the major concern of 3D tracking is speed. Our currently C++ implementation requires less than 1.5 seconds per frame, which contains $160 \times 148 \times 208 = 4,925,440$ voxels.

5. CONCLUSIONS

In this paper, we present a robust, fast and accurate LV tracking algorithm for LV in the 3D echocardiography. Instead of building specific models of the heart, all the major steps in our algorithm are based on learning. Our proposed algorithm is therefore general enough to be extended to other 3D medical tracking problems.

6. REFERENCES

[1] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu, "Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features," *ICCV*, 2007.

[2] W. Hong, B. Georgescu, X. S. Zhou, S. Krishnan, Y. Ma, and D. Comaniciu, "Database-guided simultaneous multi-slice 3D segmentation for volumeric data," *ECCV*, vol. 4, pp. 397–409, 2006.

[3] M. P. Jolly, "Automatic segmentation of the left ventricles in cardiac MR and CT images," *IJCV*, vol. 70, no. 2, pp. 151–163, 2006.

[4] Q. Duan, E. Angelini, S. Homma, and A. Laine, "Validation of optical-flow for quantification of myocardial deformations on simulated RT3D ultrasound," *ISBI*, pp. 944–947, 2007.

[5] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models: Their training and application," *CVIU*, vol. 61, no. 1, pp. 38–59, 1995.

[6] Z. Tu, "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering," *ICCV*, vol. 2, pp. 1589–1596, 2005.

[7] J. Tenebaum, V. de Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[8] F.L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *PAMI*, vol. 11, no. 6, pp. 567–585, 1989.

[9] F. Orderud, J. Hansgård, and S. I. Rabben, "Real-time tracking of the left ventricle in 3D echocardiography using a state estimation approach," *MICCAI*, vol. 4791, pp. 858–865, 2007.

[10] Y. Zhu, X. Papademetris, A. Sinusas, and J. S. Duncan, "Segmentation of myocardial volumes from real-time 3D echocardiography using an incompressibility constraint," *MICCAI*, vol. 4791, pp. 44–51, 2007.